第 39 卷第 8 期
录用定稿
计算机应用研究
Application Research of Computers
Vol. 39 No. 8
Accepted Paper

基于 Transformer 的多分支单图像去雨方法 *

谭富祥[1a, 1b, 2]，钱育蓉[1a, 1b, 2†]，孔钰婷[1a, 1b, 2]，张　昊[1a, 1b, 2]，周大新[1a, 1b, 2]，范迎迎[1a, 1b, 2]，陈　龙[1a, 1b, 2]

(1. 新疆大学 a. 软件学院; b. 软件学院重点实验室，乌鲁木齐 830046; 2. 新疆维吾尔自治区信号检测与处理重点实验室，乌鲁木齐 830046)

摘　要：雨纹会严重降低拍摄图像的质量，影响后续计算机视觉任务。为了提高雨天图像的质量，提出了一种基于 Transformer 的单图像去雨算法。首先，该算法通过具有窗口机制的 Transformer 获得大范围的感受野，进而获取雨纹特征的上下文信息，提高模型提取雨纹特征的能力；其次，该算法通过多分支模块提取和融合不同种类、不同层次的特征，提高模型对复杂雨纹信息的表征能力；最后通过残差连接融合浅层特征和深层特征，补全深层特征中缺失的细节信息，增强网络表达能力。在公开数据集 Rain100L，Rain100H 和私有数据集 Rain3000 上的实验结果表明，该方法相较于现有算法，能更有效的去除雨纹，同时更好的恢复图像中丢失的背景纹理信息。峰值信噪比和结构相似度(PSNR/SSIM)分别达到 38.33/0.9855、28.42/0.9000、34.51/0.9643。

关键词：单图像去雨；多分支；Transformer；特征融合

中图分类号：TP　　doi: 10.19734/j.issn.1001-3695.2021.12.0695

Multi-branch single image deraining network based on Transformer

Tan Fuxiang[1a, 1b, 2], Qian Yurong[1a, 1b, 2†], Kong Yuting[1a, 1b, 2], Zhang Hao[1a, 1b, 2], Zhou Daxin[1a, 1b, 2], Fan Yingying[1a, 1b, 2], Chen Long[1a, 1b, 2]

(1. *a. College of Software*, *b. Key Laboratory of Software Engineering*, *Xinjiang University*, *Urumqi* 830046, *China*; 2. *Key Laboratory of Signal Detection & Processing in Xinjiang Uygur Autonomous Region*, *Urumqi* 830046, *China*)

Abstract: Rain streaks can seriously degrade the quality of captured images and affect subsequent computer vision tasks. In order to improve the quality of rainy images, this paper proposed a single-image deraining algorithm based on Transformer. First, the algorithm obtains a wide range of receptive fields through the Transformer with window mechanism, and then obtains the contextual information of rain streak features to improve the ability of the model to extract rain streak features; secondly, the algorithm extracts and fuses different kinds and levels of features through multi-branch modules to improve the model's ability to characterize complex rain streaks information; finally, this paper fuses the shallow features and deep features through residual connections to complete the missing details in the deep features, which enhances the expression ability of the network. The experimental results on the public datasets Rain100L, Rain100H and the private dataset Rain3000 show that the method is more effective in removing rain streaks compared to existing algorithms while better recovering the lost background texture information in the images. PSNR and SSIM have respectively reached 38.33/0.9855, 28.42/0.9000 and 34.51/0.9643.

Key words: single image deraining; multi-branching; Transformer; feature fusion

0　引言

雨天作为一种常见天气，会降低所拍摄图像或视频的质量，限制图像分类，目标检测，图像分割等计算机视觉任务的应用场景。相比于视频，单图像缺少时序信息，因此研究单图像去雨更具有挑战性。

单图像去雨任务主要是依据雨纹及其周围的像素信息恢复损失背景信息，其方法大致分为传统方法和深度学习的方法[1,2]。传统方法是依据雨纹的先验知识设计模型。Chen 等人[3]根据雨纹的几何尺寸具有相似性，构建低秩表示的方法去除雨纹。Li 等人[4]从雨纹特征的稀疏性入手，使用稀疏判别字典去雨。Li 等人[5]提出高斯混合模型用于相似块补全图像的方法，实现单图像去雨。Kang 等人[6]首先将图像分解成高低频，其次采用稀疏编码处理高频信息的方法去除雨纹。虽然这些方法取得一定的效果，但在雨纹密集，复杂和背景难识别的地方，存在去雨不足或过度去雨的问题。

深度学习中基于卷积神经网络(convolutional neural networks , CNN)的方法具有强大的特征表示能力，能有效的学习从有雨图像到无雨图像的非线性映射。Fu 等人[7]提出的 DerainNet 模型首次将 CNN 方法应用到单图像去雨领域，该模型先将输入图像分为高频细节层和低频基础层，高频层用于训练去雨网络，低频层用于图像增强。Du 等人[8]认为雨纹在不同的空间位置和通道是有差异的，因此提出自适应雨纹密度的条件变分单图像去雨网络。Zhang 等人[9]同样从密度的角度考虑，构建多流密度估计器实现自适应图像去雨。He 等人[10]联合考虑雨纹密度和雨滴尺寸，提出多尺度雨纹密度估计模块指导网络去雨。Jiang 等人[11]进一步研究了多尺度模型对去雨任务的有效性，提出多尺度渐进融合模型。Wang 等

收稿日期：2021-12-13；**修回日期：**2022-02-21　**基金项目：**国家自然科学基金资助项目(61966035)；国家自然科学基金联合基金资助项目(U1803261)；自治区科技厅国际合作项目(2020E01023)；智能多模态信息处理团队项目(XJEDU2017T002)

作者简介：谭富祥(1994-)，男，新疆乌鲁木齐人，硕士研究生，主要研究方向为单图像去雨；钱育蓉(1980-)，女(满)(通信作者)，新疆乌鲁木齐人，教授，博导，博士，主要研究方向为网络计算、遥感图像处理(qyr@xju.edu.cn.)；孔钰婷(1997-)，女，湖北武穴人，硕士研究生，主要研究方向为数据挖掘；张昊(1996-)，男，山西太原人，硕士研究生，主要研究方向为小目标检测；周大新(1996-)，男，新疆乌鲁木齐人，硕士研究生，主要研究方向为低照度图像增强；范迎迎(1991-)，女，新疆乌鲁木齐人，博士，主要研究方向为遥感图像分类；陈龙(1995-)，男，山东济南人，硕士研究生，主要研究方向为超分辨率重建.

人[12]同样注意到多尺度信息对去雨任务的重要性，提出通过尺度聚合模块和自注意模块学习不同尺度的特征。

目前基于 CNN 的方法取得一定的效果，但 CNN 通过卷积层实现局部像素间接相关的方式，造成感受野有限。现有的去雨模型大都是通过堆叠卷积核扩大感受野，这种方式获得感受野仍然有限，并且会减弱特征长期依赖，造成去雨不足或过度去雨。

近期流行的 Transformer[13]具有的全局计算特性，能有效获得全局注意力图和特征长距离依赖，已被用于图像分类[14]，图像分割[15,16]等领域。但是 Transformer 不加限制的计算方式并不适合单图像去雨任务，因此，受 Swin Transformer[17]的启发,本文结合 Transformer、窗口机制以及去雨任务的特性设计了一种多分支窗口 Transformer 去雨网络(Multi-branch window Transformer network for single image deraining，MBWTNet)。该模型的特征提取模块具有感受野大，雨纹特征表达能力强的优点，多分支模块能自适应的学习不同种类，不同层次的雨纹特征，丰富特征表达。实验结果表明，本文的方法既能有效的去除复杂雨纹又能较好的恢复被雨纹遮挡的背景纹理，与目前主流的单图像去雨模型相比，获得了最佳的去雨效果。

1　Transformer 介绍

1.1　Transformer 模型介绍

Transformer 是 Vaswani 等人[13]提出用于解决自然语言处理(natural language processing, NLP)中循环神经网络不能并行处理的问题，其标准模型如图 1 所示，由左部的 Encode 和右部的 Decode 组成。在 Encode 阶段，首先将句子中的单词转换成词向量；然后通过自注意力模块，残差连接和层归一化得到全局自注意力特征图；最后通过前馈网络，残差连接和层归一化获得 Encode 的输出。与 Encode 相比，Decode 只多了一个注意力模块和归一化层用于接收 Encode 输出。Decode 的输入除了 Encode 的输出还包括上一个 Decode 的输出。Decode 输出的是对应位置的概率分布。由于并行输入缺少单词的位置关系，Transformer 使用位置编码的方式保留位置关系。

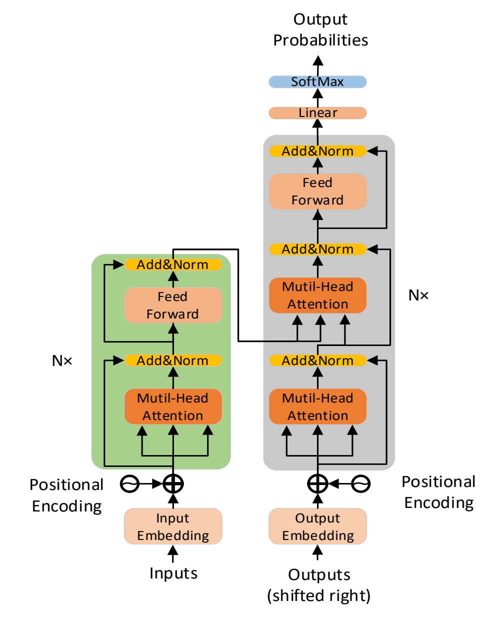


图 1　标准 Transformer 模型

Fig. 1　Standard Transformer model

Dosovitskiy 等人[14]提出的 VIT 模型是首次直接使用 Transformer 的 Ecode 部分用于图像分类，为后续视觉 Transformer 奠定了基础。为了适应 Transformer 的输入，VIT 首先将图像分割成不重叠的图像块，再将图像块拉伸并嵌入位置编码，得到一维的向量。后续视觉 Transformer 的研究大都使用这种方式输入图像或特征图。对于输出，VIT 通过分类器处理 Encode 的输出特征，得到预测结果。VIT 和 MBWTnet 都采用了相对位置编码,但不同的是 MBWTNet 在自注意力中添加位置编码。

1.2　多头自注意力机制介绍

多头自注意力是 Transformer 的重要组成部分,其结构如图 2 所示。

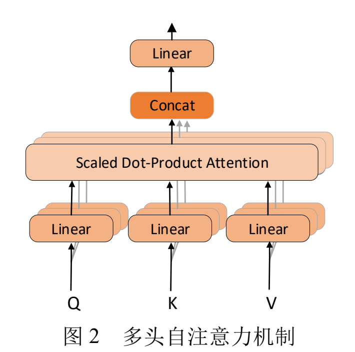


图 2　多头自注意力机制

Fig. 2　Mutil-head self-attention

首先，Encode 的输入矩阵通过 3 个不同权重的变换矩阵得到的查询矩阵 Q，键矩阵 K 和值矩阵 V。然后通过点积注意力，如表达式(1)，计算自注意力特征图；多头自注意力是通过多组变换矩阵和等式(1)得到多个相互独立的注意力特征图。最后通过拼接和全连接融合不同的注意力特征图得到多头注意力图。

$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

式(1)中，Q，K，V 是向量组成的二维矩阵，Q 与 K 转置的点乘得到的相关性矩阵记录了所有向量之间的相关性，而 Q 和 K 来自同一个矩阵的变换，因此，相关性矩阵描述的是输入向量之间的相关性。为了避免 $\mathrm{softmax}(\cdot)$ 造成梯度消失，使用一个系数等效缩放相关性矩阵。经过激活的相关性矩阵与 V 点乘得到全局自注意力图。多头自注意力是 Transformer 的全局感受野和特征长距离依赖的主要来源。

2　多分支窗口 Transformer 去雨网络

Transformer 的全局计算的方式使模型具有全局感受野和特征长期依赖，但会造成一定的特征冗余，不适合直接用于单图像去雨。本文提出一种多分支窗口 Transformer 去雨网络 MBWTNet，该网络模型通过窗口限制计算获得较大的感受野，充分利用 Transformer 与多分支结合的优势以及残差连接提取不同层次的特征。如图 3 所示，MBWTNet 由基于 Transformer 的特征提取模块(Transformer-based feature extraction block，TFEB)、多分支特征融合模块(Multi-branch fusion module，MBFM)和残差连接构成。在 CNN 中，残差连接是为了解决较深网络中梯度消失的问题，本文中残差连接更关注浅层特征的作用，即补全深度特征中缺失的纹理信息。MBWTNet 采用三个顺序排列的 MBFM 模块提取和融合不同层次的特征，其中，前两个 MBFM 的输出以残差方式传递到网络深层，实现浅层特征与深层特征的充分融合。第三个 MBFM 的输出被输入到三个并列的 TFEB，通过增加网络的深度和宽度，同时提取不同种类的特征。网络的计算过程如式(2)所示。

$$\begin{cases} x_t = F_{MBFM}^t(x_{t-1}), t=1,2,3 \\ x_4 = TFEB_4(x_2 + TFEB_1(x_3) + TFEB_2(x_3) + TFEB_3(x_3)) \\ x_{pre} = TFEB_5(x_1 + x_4) \end{cases} \tag{2}$$

式(2)中，$F_{MBFM}^t(\cdot)$ 是多分支特征融合模块，$\mathrm{TFEB}_i(\cdot)$ 是特征

提取模块，x_i 是中间变量，拥有相同的尺寸和通道，其中 x_0 为输入的有雨图像，x_{pre} 为预测图像。

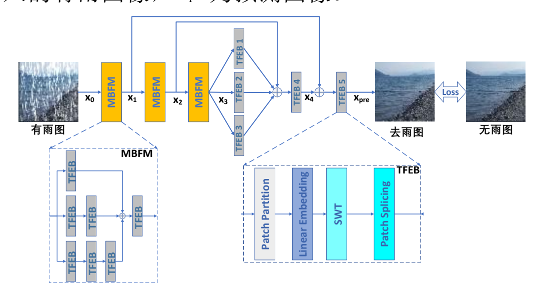


图 3　基于 Transformer 的多分支去雨网络模型

Fig. 3　Transformer-based multi-branch deraining network

2.1　特征提取模块 TFEB

由于卷积操作无法充分获得像素之间的特征联系造成一些基于 CNN 的方法去除长条状雨纹的效果不理想，即模型存在去雨不足或过度去雨。Transformer 的全局计算方式能充分获得像素之间的联系，但会造成特征冗余。针对该问题，本文在 swin Transformer[17]的基础上构建了一个特征提取模块。该模块采用 swin Transformer 的窗口滑动机制限制计算量和实现窗口间信息的交流。swin Tansformer 会造成一定空间信息的损失，针对该问题，本文提出一个图像块拼接模块(Patch Splicing)避免空间信息损失。特征提取模块如图 3 中 TFEB 部分所示，特征图依次通过分割模块(patch partition)，维度调整模块(linear embedding)，基于滑动窗口的 Transformer 模块(swin-Transformer block，SWT)和图像块拼接模块(Patch Splicing)，完成特征的提取。

patch partition 先将输入尺寸为 H×W×C 的特征图分割成不重叠的 Patch 块，如式(3)所示，其中 H 和 W 是输入图像尺寸，Wp 和 Hp 为 Patch 块的尺寸。

$$(P_1,P_2,P_3,\cdots P_n)=\mathrm{F_{split}}(x_{in}),\mathrm{n}=\frac{\mathrm{w}}{\mathrm{w_p}}\times\frac{\mathrm{H}}{\mathrm{H_p}} \tag{3}$$

$$\hat{P}_t=\mathrm{F_{reshapea}}(P_t),\ \ \mathrm{t}=1,2,3,\cdots,\mathrm{n} \tag{4}$$

由于 Transformer 只接受一维向量。Patch Partition 再通过 $\mathrm{F}_{reshapea}(\bullet)$ 将 Patch 块 $P\in R^{\mathrm{w_p\times H_p\times c}}$ 按通道方向转换成 1D 的向量 $\hat{P}\in R^{(\mathrm{w_p\times H_p\times c})}$，该向量可以视为一个“token”。Patch 块的尺寸与位置编码紧密相关，Patch 块的尺寸越大，位置编码的尺寸越小。图像分类等其他计算机视觉任务中更多关注的是语义信息，例如图像分类模型 VIT[14]和图像分割[16]的 patch 块都设为 16×16，位置编码的尺寸为 $\frac{\mathrm{H}}{16}\times\frac{\mathrm{W}}{16}$，而在图像去雨任务中更多关注的是像素信息和位置信息。因此，本文中 patch 块的尺寸为 3×3，即 $\mathrm{w_p}=3$，$\mathrm{H_p}=3$。

经过 Patch Partition 分割后的向量维度为 $3\times3\times\mathrm{C}$，考虑到高维具有更高的特征表达能力，有利于自注意力模块学习雨纹特征，本文在维度调整模块 Linear Embedding 中通过全连接 $\mathrm{F}_{lineara}(\bullet)$ 将向量的维度映射到 $3\times3\times\mathrm{C}\times2$，即式(5)中 $Z\in R^{(3\times3\times\mathrm{C}\times2)}$。

$$Z=\mathrm{F_{lineara}}(\hat{P}) \tag{5}$$

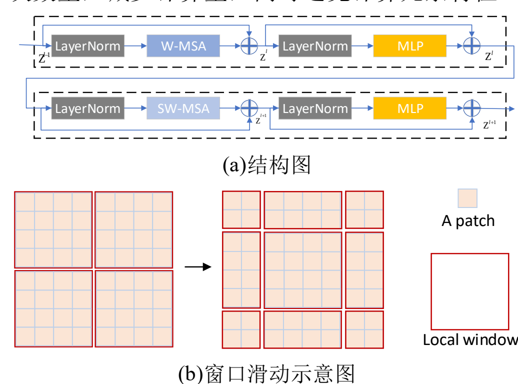
标准的 Transformer 具有全局关注，特征远距离依赖的优点，但存在计算量大，模型部署难的问题。受文献[17]启发，本文采用滑动窗口的方式限制计算，模型结构如图 4(a)所示，每个子模块由两个 LayerNorm(LN)层，一个基于 7×7 窗口的多头自注意力模块(W-MSA)和一个 MLP 构成，其中多头注意力的头数为 3。patch 块包含 3×3 个像素，因此 7×7 窗口的感受野为 21×21。相比于卷积层，基于窗口的 Transformer 能获得较大的感受野，进而更充分的提取窗口内不同尺寸的特征。由于窗口边界缺少足够的纹理信息用于特征提取，受 swin Transformer[17]的启发，该模块的第二个子模块采用滑动的多头自注意模块(SW-MSA)，即窗口位置与第一个不同，如图 4(b)所示 。滑动窗口机制使边界像素信息在同一个窗口内，完成边界雨纹特征的学习和窗口间信息交流。基于滑动窗口的 Transforme 模型计算过程如式(6)所示。

$$\begin{cases} \hat{Z}^{l}=\mathrm{W-MSA}(\mathrm{LN}(Z^{l-1}))+Z^{l-1} \\ Z^{l}=\mathrm{MLP}\left(\mathrm{LN}\left(\hat{Z}^{l}\right)\right)+\hat{Z}^{l} \\ \hat{Z}^{l+1}=\mathrm{SW-MSA}(\mathrm{LN}(Z^{l}))+Z^{l} \\ Z^{l+1}=\mathrm{MLP}\left(\mathrm{LN}\left(\hat{Z}^{l+1}\right)\right)+\hat{Z}^{l+1} \end{cases} \tag{6}$$

式(6)中，$\hat{Z}^{l}\in R^{\frac{\mathrm{H}}{3}\times\frac{\mathrm{W}}{3}\times(3\times3\times2\times\mathrm{C})}$ 表示多头注意力的输出，$Z^{l}\in R^{\frac{\mathrm{H}}{3}\times\frac{\mathrm{W}}{3}\times(3\times3\times2\times\mathrm{C})}$ 表示全连接的输出。与标准的 Transformer 中多头注意力模块不同，W-MSA 使用的注意力模块如式(7)所示。

$$Attention(Q,K,V)=SoftMax\left(QK^{T}/\sqrt{d}+B\right)V \tag{7}$$

式(7)中，$Q,K,V\in R^{\mathrm{M^2\times d}}$，$Q$ 代表查询(Query)，K 代表关键字(Key)，V 代表值(Veal)，M^2 是参与计算的 patch 块数量(7×7)，d =32 是 Query、Key 或 Veal 的维度，$B\in\mathbb{R}^{M^2\times M^2}$ 是位置编码。W-MSA 通过限制单次参与 self-attention 计算的 patch 块数量，减少计算量，同时避免计算冗余特征。



(a)结构图

(b)窗口滑动示意图

图 4　基于滑动窗口的 Transformer 模块(SWT)

Fig. 4　Sliding window-based Transformer block (SWT)

输出尺寸与输入尺寸一致是特征提取模块可直接用于模块的堆叠的必要条件，也有利于融合不同层次的特征。图像块拼接模块 patch splicing 首先通过全连接 $\mathrm{F_{linearb}}(\cdot)$ 将经过 Transformer 计算的高维向量 f_{in} 的维度压缩至 $3\times3\times\mathrm{C}$。这不仅考虑了 patch splicing 模块的输入是高维向量，不符合构建 patch 块的要求，也考虑了全连接能自适应的保留重要特征，抑制次要特征。然后通过 $\mathrm{F_{reshapeb}}(\bullet)$ 按通道方向将向量转换成 patch 块，即 $s_r\in\mathrm{R}^{3\times3\times\mathrm{C}}$。最后将这些 patch 块拼接成特征图 $f_{out}\in R^{\mathrm{W\times H\times C}}$。上述计算过程如式(8)所示，其中 $s_l\in R^{(3\times3\times\mathrm{C})}$。

$$\begin{cases} s_l=F_{linearb}(f_{in}) \\ s_r=F_{reshapeb}(s_l) \\ f_{out}=F_{splic}(s_r) \end{cases} \tag{8}$$

2.2　多分支特征融合模块

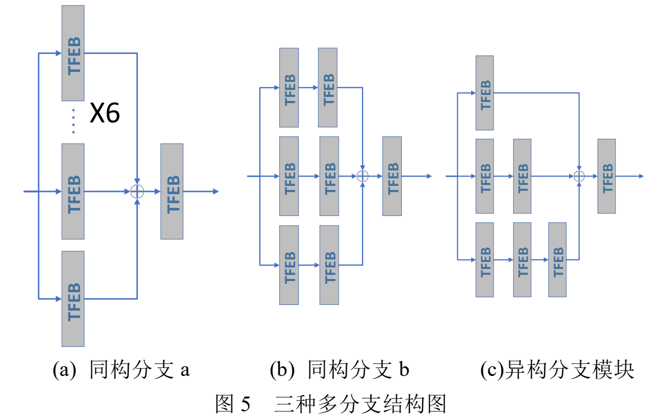
雨纹图像包含雨纹尺寸、形状等不同种类特征，背景图包含不同层次的特征。多头注意力利用网络不同初始值学习提取和融合不同种类的特征，但该方法无法学习提取和融合不同层次的特征。为了更好地满足去雨任务中特征多样性的需求，本文通过研究融合多个 TFEB 模块，提升模型去雨的性能，因此本文设计和讨论了三种多分支结构，如图 5 所示。

图 5(a)是同构多分支结构。由于各分支网络初始值不同，且相互独立，训练时向着不同的特征子空间学习。因此分支数越多，提取的特征越丰富，去雨性能越好，但是分支数量越多，并不意味着网络越好，2.4 节的多分支实验证明了这个

观点。图 5(b)拥有和图 5(a)同等的参数量，每个分支采用相同的结构，但图 5(b)的分支数只有图 5(a)的一半。图 5(c)与图 5(b)拥有相同的分支数，相同的参数量，但图 5(c)的每个分支采用不同的结构，该结构计算过程如式(9)所示。

$$f_{MBFM} = TFEB_0(TFEB_1^1(x) + TFEB_1^2(TFEB_2^2(x)) + TFEB_1^3(TFEB_2^3(TFEB_3^3(x)))) \tag{9}$$

式(9)中，$TFEB_i(\cdot)$ 表示特征提取模块。x 表示多分支模块 MBFM 的输入，f_{MBFM} 是输出。由于每个分支具有不同的初始值和结构，导致模块能自适应的学习不同种类，不同层次的特征，丰富输出的特征。通过特征相加的方式并不能充分融合不同分支的特征，本文通过添加一个特征提取模块，实现特征充分融合。



(a) 同构分支 a　　(b) 同构分支 b　　(c)异构分支模块

图 5　三种多分支结构图

Fig. 5　Three kinds of multi-branch structure diagram

2.3　损失函数

现有图像去雨模型的损失函数大多数使用的是已被 Ren 等[18]证明有效的结构相似度(SSIM structural similarity)，该损失函数虽然能获得较好的结构相似度，但生成图像的颜色存在一定程度的失真，造成峰值信噪比(PSNR)较低。在这项工作中，本文使用的损失函数的数学表达式如式(10)所示。

$$\begin{cases} Loss_{L_1} = L_1(SDnet(O), B) \\ Loss_{SSIM} = 1 - SSIM(SDnet(O), B) \\ Loss_{ide} = L1(SDnet(B), B) \\ Loss_{all} = \alpha \times Loss_{L_1} + \beta \times Loss_{SSIM} + \lambda \times Loss_{Ide} \end{cases} \tag{10}$$

式(10)中，α=0.2，β=4，λ=1，O 是有雨图像，B 是对应的背景图。 是绝对偏差和(Sum of Absolute Difference，SAD)，是基于两张图像的像素差计算的。结构相似度(Structural Similarity，SSIM)是评价两张图像内容的结构相似性的指标，其负数常被用作损失函数，表达式如式(10)中 LossSSIM 所示。身份损失(identity loss，ide)是源于 CycleGAN [19]中用于约束生成图像的颜色损失，本文将其用于约束去雨后图像的颜色差异，表达式如式(10)中 Losside 所示，将背景图作为模型的输入，生成的结果与标签通过 L1 计算身份损失。本文通过最小化三种损失值的和，使模型保持图像结构信息的同时，减小颜色差异，提高模型去雨性能。

3　实验

3.1　数据集

现有的公开数据集 Rain100L 和 Rain100H[20]是由 1800 对训练集和 200 对测试集组成的数据集，它们是在相同的背景图像上添加不同方向的雨纹。Rain100L 是去雨相对简单的数据集，每张图片包含有 1 种方向的雨纹。Rain100H 是去雨相对困难的数据集，每张图片包含 5 种方向的雨纹。Rain100L 和 Rain100H 提供两种难度的数据集用于评估网络的性能。但这两个数据集都存在训练集和测试集背景相似的问题[18]，这会降低模型的可信度。

针对这个问题，文献[18]通过剔除 546 张相似的背景，以此提高数据集的质量。但这会降低样本量，不利于模型的泛化。本文使用完全的 Rain100H 和 Rain100L 训练和测试模型，公平的对比现有的主流模型。此外提出一个全新的数据集用于提高模型的可信度。该数据集首先从内容丰富的 ImageNet 中随机选取 10 万张图片；然后从具有 825 张雨纹图的 Efficientderain[21]中随机选取 1 至 4 种雨纹添加到选取的图片中，最后从 10 万对数据集中选取前 3400 对合成的图像作为数据集，其中训练集 3000 对，测试集 400 对。本文将该数据集命名为 Rain3000，如图 6 所示。Rain3000 既包含简单的雨纹，也包含相对复杂的雨纹，这有利于拟合真实雨图的特征分布。数据集的参数如表 1 所示。

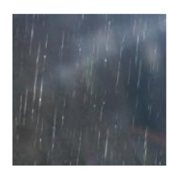
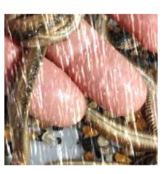
   

图 6　数据集 Rain3000

Fig. 6　Data set Rain3000

表 1　数据集对比

Tab. 1　Comparison of data sets

数据集名称	训练集	测试集	图像尺寸	雨纹种类
Rain100L	1800	200	321×481	1
Rain100H	1800	200	321×481	5
Rain3000	3000	400	256×256	1-4

为了验证所提数据集训练网络的有效性，本文首先通过在数据集 Rain3000、Rain100L、Rain100H 分别训练 DCSFN[22]、MPRnet[23]和 PRENet[24]，然后在真实雨图上测试，结果如图 7 所示。通过 Rain3000 进行训练，DCSFN 模型能很好的去除不同形状，大小的雨纹，保留背景信息，MPRNet 和 PRENet 能去除较小，更接近自然的雨纹，这说明数据集 Rain3000 能更好的拟合自然界雨纹特征的分布。

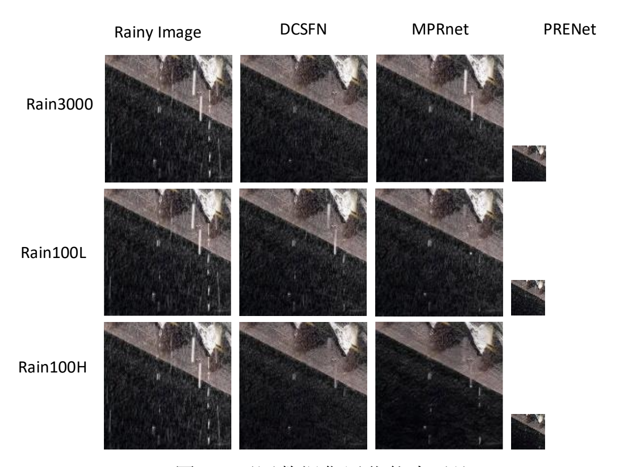


图 7　不同数据集泛化能力对比

Fig. 7　Comparison of generalization ability of different datasets

3.2　实验设置

本文实验环境为 GPU 显卡 Tesla V100 16G，内存 32GB，使用 pytorch 深度学习框架，版本号 Pytorch 1.7.0，batch size 设置为 5，总共训练 500 个 epoch。学习率的初始值为 5×10^{-4}，分别在总迭代次数的 3/5 和 4/5 时衰减为 5×10^{-5} 和 5×10^{-6}。本文在据集 Rain100L，Rain100H 和 Rain3000 上对比主流算法，在数据集 Rain3000 上进行消融实验。

本文使用已被广泛用于评估去雨性能结构相似性(SSIM)和峰值信噪比(PSNR)。SSIM 是度量两张图像内容，纹理的相似性指标。SSIM 最大值是 1，越接近于 1，表示两张图片的越相似。PSNR 是基于两张图片之间的像素误差计算的，误差越小，值越大，图片越相似，去雨的效果越好，反之图

像去雨的效果越差。

3.3 对比实验

为了验证 MBWTNet 的优越性，本文在数据集 Rian100L，Rain100H 和 Rain3000 上对比了如下 6 种先进的去雨方法：

a) RESCAN：recurrent squeeze-and-excitation context aggregation net method[25] (ECCV，2018)，使用递归结构分多个阶段去雨，每个阶段使用多个具有 SE(Squeeze-and-Excitation)模块和膨胀卷积的上下文聚合网络，此外，该网络还设计了一个记忆单元用于增强不同阶段之间的联系。

b) GCANet：gated context aggregation network [26](WACM，2019)，提出一种使用平滑扩张卷积的上下文聚合网络用于去雾，解决了因膨胀卷积引起的栅格化。该方法同样适用图像去雨。

c) NLEDN：non-locally enhanced encoder-decoder network[27] (ACMMM，2018)，该方法提出非局部增强自编码网络使用区域级非局部增强，提高捕获空间上下文远程依赖关系的能力，此外使用串连不同尺度区域的方式增强区域间交流。

d) PREnet：progressive image deraining network method[24] (CVPR，2019)，提出一个多阶段去雨的基线模型，每个阶段的输入是原始雨图和上个阶段输出的拼接，此外，还使用一个 LSTM 挖掘不同阶段之间的深层特征。

e) DCSFN：deep cross-scale fusion net-work for single image rain removal[22] (ACMMM，2020)，提出一种跨尺度融合方法来学习不同尺度之间的内部特征联系，此外，使用密集连接增强远程空间依赖性。

f) MPRnet：multi-stage progressive image restoration[23] (CVPR，2021)，提出一种多阶段渐进修复模型用于平衡修复图像时空间细节和上下文信息，每个阶段都使用标签进行监督，此外，其夸阶段聚合多尺度特征的策略实现不同阶段间信息交换。

表 2　与其他算法对比结果

Tab. 2　Comparison results with other algorithms

方法	Rain3000		Rain100L		Rain100H	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
RESCAN	0.9248	30.75	0.9629	33.99	0.7612	23.98
NLEDN	0.9554	33.44	0.9806	37.2	0.8654	27.21
GCANet	0.9354	31.86	0.976	36.19	0.8203	25.99
PREnet	0.9547	32.81	0.9787	35.94	0.8661	26.1
DCSFN	<u>0.9539</u>	33.33	<u>0.9821</u>	37.603	<u>0.886</u>	<u>27.76</u>
MPRnet	0.9531	<u>33.83</u>	0.979	<u>37.81</u>	0.8484	26.55
MBWTNet	**0.9643**	**34.51**	**0.9853**	**38.33**	**0.9000**	**28.42**

表 2 中评价指标最优值用加粗表示，次优值用下划线表示。分析结果可知，本文的算法在数据集 Rain3000、Rain100L 和 Rain100H 上均获得最好的性能。在去雨难度相对简单的 Rain100L 数据集上，PSNR 能达到 38.33dB。在去雨任务困难的 Rain100H 数据集上，PSNR 能达到 28.42dB。在雨纹特征复杂的 Rain3000 上，PSNR 达到 34.51dB。本文的算法在数据集 Rain100H 上的优势最明显，相比于 2018 年的 RESCAN 网络，评价指标 PSNR 和 SSIM 分别提升 4.44dB，0.1388，相比于最新的 MPRNet 模型，PSNR 和 SSIM 分别提升 1.87dB，0.0516，相比于次优的 DCSFN，PSNR 和 SSIM 在分别提升 0.66dB，0.014。这表明，相比于 RESCAN 和 GCAN 使用膨胀卷积获得的感受野，MBWTNet 拥有更广阔的感受野，更强的特征表示能力；相比于 PREnet，DCSFN 和 MPREnet 增强特征依赖的方式，MBWTNet 拥有更强的特征长距离依赖，更丰富的特征表达；相比于 NLEDN 使用多尺度实现增强区域间信息交流，MBWTNet 的滑动窗口方式具有更充分，更直接的优点。

图 8 展示了各个算法去雨的视觉效果。可以看出，RESCAN 去雨后的图像存在伪影，NLEDN，CGAN，DCSFN，MPRNet 虽然取得了较好的去雨效果，但仍然有一些长条状的雨纹未去除。PREnet 虽然能去除雨纹，但也去除了背景中一些纹理细节。这六种模型在恢复效果上都存在一定的不足，而 MBWTNet 既能很好的去除雨纹又能较为满意的恢复纹理细节，这进一步证明了所提方法的优越性。



(a)有雨图　(b)RESCAN　(c)NLEDN

(d)CGAN　(e)PRENet　(f)DCSFN

(g)MPRNet　(h) MBWTNet　(i)无雨图

图 8　不同算法在数据集 Rain3000 上的去雨效果

Fig. 8　Deraining effect of different algorithms on dataset Rain3000

模型参数量和预测时间是模型实用性的重要指标，图 9 展示了各个模型的参数量和实时性，从图中可以看出，虽然所提模型的参数量较大，但却获得了最快的推理速度。这是因为基于滑动窗口的 Transformer 和全连接采用了矩阵运算的方式，这比逐步卷积的方式要高效。图 9 也进一说明所提算法的实用性。

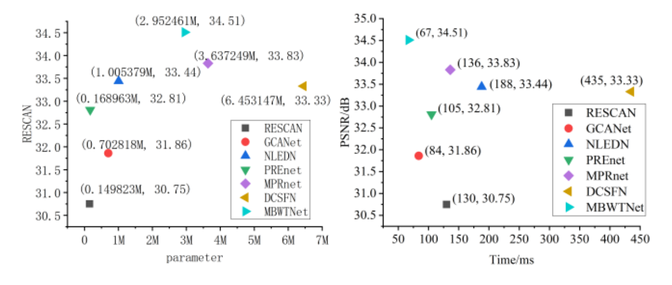


图 9　模型的参数量和实时性对比

Fig. 9　Comparison of the number of parameters and the real-time performance of the model

3.4 消融实验

3.4.1 分支数量及结构对去雨性能的影响

为了证明分支的数量和结构对模型去雨性能的影响，本文在数据集 Rain3000 上做了两组对照实验。第一组对照实验是验证分支结构相同，分支数量对模型性能的影响，模型的其他部分不改变，只将多分支融合模块 MBFM 替换成如图 5(a)所示的结构，更改分支的数量为 1,2,3,4,5,6 进行实验，实验结果如图 10 所示。从结果可以看出，随着分支数增多，模型去雨的性能越好，但分支数超过 4 之后，模型性能提升有限。这是因为相同结够的分支数量越多，分支提取的特征越相似，这限制特征多样性的进一步表达。本文为了平衡模型

的规模和性能，本文采用三分支结构。

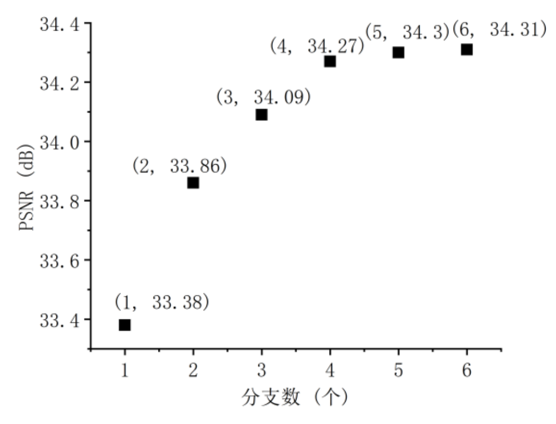


图 10 分支数实验结果图

Fig. 10　Graph of experimental results of branching number

第二组对照实验是验证参数量相同时，相同结构的分支与不同结构的分支对网络性能的影响，模型的其他部分不改变，只改变多分支融合模块 MBFM 的结构为图 5(a)(b)和(c)，实验结果如表 3 所示，表中 MBFM-a 对应图(a)，其他的依此类推。

表 3　分支结构实验结果

Tab. 3　Experimental results of branching structure

结构名称	SSIM	PSNR
MBFM-a	0.9636	34.31
MBFM-b	0.9640	34.44
MBFM-c	0.9643	34.51

分析表 3，在分支数和参数量相同时，分支结构不相同的模块具有更丰富的特征表达能力。结合图 10 和表 3 可知，相同分支数量的增加虽然能提升模型去雨的效果，但这种提升效果有限，不同结构的分支能极大程度的自适应捕获不同种类，不同层次特征之间的相关性。因此，避免分支结构相同，更有利于学习从有雨图像到无雨图像的映射。

3.4.2 损失函数的性能

为了验证不同损失函数对模型去雨性能的影响，分别对本文所涉及的损失函数进行了实验。模型采用相同结构的三分支，评价指标采用 SSIM 和 PSNR。表 4 展示了使用不同的损失函数训练网络得到的性能。通过分析可得，将 Losssim 作为损失函数时，网络能获得最好的 SSIM 指标。将 Lossssim 和 Losside 作为损失函数时能获得与 Lossssim 作为损失函数时相同的 SSIM 指标，并使 PSNR 提升了 0.03dB。而添加 Losside 会降低 LossL1 的性能，因为这两个损失函数都是基于像素差值计算的，Losside 抑制了 LossL1 的性能。另外 LossL1 也会限制仅有 Lossssim 时的性能，只有在三种损失函数都包含时，网络才能获得最佳性能，相比于仅有 Lossssim，PSNR 提升了 0.07dB。因为 Lossssim 是基于图像内容结构计算的损失值，缺少像素纹理等信息，而 LossL1 和 Losside 能有效的从不同的角度补充像素信息，也说明用于约束图像风格转换任务中，颜色差异的身份损失同样能约束图像去雨任务中颜色的差异。

表 4　损失函数的对比

Tab. 4　Comparison of loss functions

Loss-Function	Lossssim	LossL1	Losside	SSIM	PSNR/dB
SDNet-ssim	✓	-	-	0.9620	34.02
SDNet-ssim_ide	✓	-	✓	0.9620	34.05
SDNet-L1	-	✓	-	0.9587	33.97
SDNet-L1_ide	-	✓	✓	0.9551	33.52
SDNet-ssim_L1	✓	✓	-	0.9619	34.01
SDNet-ssim_L1_ide	✓	✓	✓	0.9620	34.09

4　结束语

针对图像去雨，本文提出一种多分支窗口 Transformer 去雨网络(MBWTNet)，该网络首先结合 Transformer 和窗口机制构建一种局部像素直接相关，大范围感受野和无空间信息损失的特征提取模块；然后基于该模块构建了一种多分支模块用于提取和融合不同种类、不同层次的特征；最后实用前馈网络和跳跃连接构建端到端的去雨网络。此外本文提出一个基于 ImageNet 制作的去雨数据集 Rain3000，该数据集由 3000 对训练集和 400 对测试集组成，具有背景纹理丰富，雨纹种类多样的优点。本文提出的模型在公开数据集 Rain100L，Rain100H 和私有数据集 Rain3000 上对比了几种深度学习方法，在视觉观感和定量指标上都取得了最好的结果，但存在一定局限性，例如，算法中缺少对通道相关性的描述，进一步的研究将考虑结合全局通道注意力和窗口通道注意力，提升模型捕获通道相关性的能力。

参考文献：

[1] 张育龙，王强，陈明康，等. 图像去雨算法在云物联网应用中的研究综述 [J]. 计算机科学, 2021, 48 (12): 231-242. (Zhang Yulong, Wang Qiang, Chen Mingkang, *et al.* Survey of intelligent rain removal algorithms for cloud-iot systems [J]. Computer Science, 2021, 48 (12): 231-242.)

[2] 陈舒曼，陈玮，尹钟. 单幅图像去雨算法研究现状及展望 [J]. 计算机应用研究, 2022, 39 (1): 9-17. (Chen Shuman, Chen Wei, Yin Zhong. Research status and prospect of single image rain removal algorithm [J]. Application Research of Computers, 2022, 39 (1): 9-17.)

[3] Chen Yilei, Hsu Chiouting. A generalized low-rank appearance model for spatio-temporally correlated rain streaks [C]// Proc of the IEEE International Conference on Computer Vision. 2013: 1968-1975.

[4] Li Yu, Tan R T, Guo Xiaojie, *et al.* Rain streak removal using layer priors [C]// Proc of the IEEE conference on computer vision and pattern recognition. 2016: 2736-2744.

[5] Li Siyuan, Ren Wenqi, Zhang Jiawan, *et al.* Single image rain removal via a deep decomposition–composition network [J]. Computer Vision and Image Understanding, 2019, 186: 48-57.

[6] Kang Liwei, Lin Chiawen, Fu Yuhsiang. Automatic single-image-based rain streaks removal via image decomposition [J]. IEEE Trans on image processing, 2011, 21 (4): 1742-1755.

[7] Fu Xueyang, Huang Jiabin, Ding Xinghao, *et al.* Clearing the skies: A deep network architecture for single-image rain removal [J]. IEEE Trans on Image Processing, 2017, 26 (6): 2944-2956.

[8] Du Yingjun, Xu Jun, Zhen Xiantong, *et al.* Conditional variational image deraining [J]. IEEE Trans on Image Processing, 2020, 29: 6288-6301.

[9] Zhang He, Patel V M. Density-aware single image de-raining using a multi-stream dense network [C]// Proc of the IEEE conference on computer vision and pattern recognition. 2018: 695-704.

[10] He Jingwei, Lei Yu, Xia Guisong, *et al.* Single image deraining with continuous rain density estimation [J]. IEEE Trans on Multimedia, 2021.

[11] Jiang Kui, Wang Zhongyuan, Yi Peng, *et al.* Multi-scale progressive fusion network for single image deraining [C]// Proc of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8346-8355.

[12] Wang Hong, Xie Qi, Zhao Qian, *et al.* A model-driven deep neural network for single image rain removal [C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3103-3112.

[13] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need [C]// Advances in neural information processing systems. 2017: 5998-6008.

[14] Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale [EB/OL]. (2021-06-03) . https://arxiv. org/abs/2010. 11929.

[15] Chen Jieneng, Lu Yongyi, Yu Qihang, *et al.* Transunet: Transformers make strong encoders for medical image segmentation [EB/OL]. (2021-02-08) . https://arxiv. org/abs/2102. 04306.

[16] Ranftl R, Bochkovskiy A, Koltun V. Vision Transformers for dense prediction [C]// Proc of the IEEE/CVF International Conference on Computer Vision. 2021: 12179-12188.

[17] Liu Ze, Lin Yutong, Cao Yue, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows [EB/OL]. (2021-08-17) . https://arxiv. org/abs/2103. 14030.

[18] Ren Dongwei, Zuo Wangmeng, Hu Qinghua, *et al.* Progressive image deraining networks: A better and simpler baseline [C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3937-3946.

[19] Zhu Junyan, Park Taesung, Isola Phillip, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proc of the IEEE international conference on computer vision. 2017: 2223-2232.

[20] Yang Wenhan, Tan Robby T, Feng Jiashi, *et al.* Deep joint rain detection and removal from a single image [C]// Proc of the IEEE conference on computer vision and pattern recognition. 2017: 1357-1366.

[21] Guo Qing, Sun Jingyang, Felix Juefei-Xu, *et al.* Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining [EB/OL]. (2020-09-19) . https://arxiv. org/abs/2009. 09238.

[22] Wang Cong, Xing Xiaoying, Su Zhixun, *et al.* DCSFN: deep cross-scale fusion network for single image rain removal [C]// Proc of the 28th ACM international conference on multimedia. 2020: 1643-1651.

[23] Zamir S W, Arora A, Khan S, *et al.* Multi-stage progressive image restoration [C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14821-14831.

[24] Ren Dongwei, Zuo Wangmeng, Hu Qinghua, *et al.* Progressive image deraining networks: A better and simpler baseline [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3937-3946

[25] Li Xia, Wu Jianlong, Lin Zhouchen, *et al.* Recurrent squeeze-and-excitation context aggregation net for single image deraining [C]// Proc of the European Conference on Computer Vision (ECCV) . 2018: 254-269.

[26] Chen Dongdong, He Mingming, Fan Qingnan, *et al.* Gated context aggregation network for image dehazing and deraining [C]// 2019 IEEE winter conference on applications of computer vision (WACV) . IEEE, 2019: 1375-1383.

[27] Li Guanbin, He Xiang, Zhang Wei, *et al.* Non-locally enhanced encoder-decoder network for single image de-raining [C]// Proc of the 26th ACM international conference on Multimedia. 2018: 1056-1064.